

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory

Kiriliouk, Anna; Naveau, Philippe

Published in:
Annals of Applied Statistics

DOI:
[10.1214/20-AOAS1355](https://doi.org/10.1214/20-AOAS1355)

Publication date:
2020

Document Version
Peer reviewed version

[Link to publication](#)

Citation for pulished version (HARVARD):
Kiriliouk, A & Naveau, P 2020, 'Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory', *Annals of Applied Statistics*, vol. 14, no. 3, pp. 1342-1358.
<https://doi.org/10.1214/20-AOAS1355>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CLIMATE EXTREME EVENT ATTRIBUTION USING MULTIVARIATE PEAKS-OVER-THRESHOLDS MODELING AND COUNTERFACTUAL THEORY

BY ANNA KIRILIOUK^{‡,*} AND PHILIPPE NAVEAU[†]

Université de Namur and CNRS (Gif-sur-Yvette)[†]*

Numerical climate models are complex and combine a large number of physical processes. They are key tools in quantifying the relative contribution of potential anthropogenic causes (e.g., the current increase in greenhouse gases) on high impact atmospheric variables like heavy rainfall. These so-called climate extreme event attribution problems are particularly challenging in a multivariate context, that is, when the atmospheric variables are measured on a possibly high-dimensional grid.

In this paper, we leverage two statistical theories to assess causality in the context of multivariate extreme event attribution. As we consider an event to be extreme when at least one of the components of the vector of interest is large, extreme-value theory justifies, in an asymptotical sense, a multivariate generalized Pareto distribution to model joint extremes. Under this class of distributions, we derive and study probabilities of necessary and sufficient causation as defined by the counterfactual theory of Pearl. To increase causal evidence, we propose a dimension reduction strategy based on the optimal linear projection that maximizes such causation probabilities. Our approach is tested on simulated examples and applied to weekly winter maxima precipitation outputs of the French CNRM from the recent CMIP6 experiment.

1. Introduction. Quantifying human influence on climate change and identifying potential causes of climate extremes is often referred to as extreme event attribution (EEA), which falls within the research field of detection and attribution (D&A) (see, e.g. the report of [National Academies of Sciences, Engineering and Medicine, 2016](#); [Chen et al., 2018](#)). In such studies, the main inferential objective is estimation of extreme quantiles (also called return levels), well-used in finance, hydrology and other fields of risk analysis (e.g. [Embrechts, Klüppelberg and Mikosch, 1997](#)). In EEA, return

[‡]Part of this work was supported by the French national program FRAISE-LEFE/INSU, MELODY-ANR, Eupheme, FUSIMET (PEPS I3A), the DAMOCLES-COST-ACTION on compound events and the Fonds de la Recherche Scientifique - FNRS (grant number J.0124.20).

Keywords and phrases: multivariate generalized Pareto distribution, necessary and sufficient causation, heavy rainfall, climate change

levels are computed under two scenarios that differ according to the causal link of interest, often increases in greenhouse gases (GHG) concentrations (see, e.g. [Angélil, Stone and Wehner, 2017](#); [Stott et al., 2016](#); [Fischer and Knutti, 2015](#)). Typically, such an approach compares probabilities under a factual scenario of conditions that occurred around the time of the event against probabilities under a counterfactual scenario in which anthropogenic emissions never occurred. More specifically, one compares the probability of an extreme event in the factual world, denoted p_1 , to the probability of an extreme event p_0 in a counterfactual world, i.e., a world that might have been if no anthropogenic forcing would have been present. The definition of the so-called extreme event is by itself a non-trivial task and depends on the application at hand. A common choice is to take some climatological index exceeding a high threshold. In their seminal paper, [Stott, Stone and Allen \(2004\)](#) studied mean June–August temperatures in Europe in order to quantify by how much human activities may have increased the risk of European heatwaves. In this example, a one dimensional sample mean, say X , summarized a complex random field that varied in time and space. The set $\{X > v\}$ where $v = 1.6$ Kelvin was chosen to resemble the 2003 mean European summer anomaly temperatures. The probabilities p_0 and p_1 were then inferred from numerical counterfactual and factual runs respectively, using nonparametric inference techniques (for bootstrap counting methods in EEA, see [Paciorek, Stone and Wehner, 2018](#)) and univariate extreme-value theory (EVT); see, e.g. [Coles \(2001\)](#) for an introduction.

In other environmental research areas, complex multivariate EVT models are commonly used (see, e.g. [Davison and Huser, 2015](#); [Cooley, Hunter and Smith, 2017](#); [de Fondeville and Davison, 2018](#); [Engelke, De Fondeville and Oesting, 2018](#); [Reich, Shaby and Cooley, 2014](#); [Shaby and Reich, 2012](#)). Bayesian hierarchical modeling (see, e.g. [Hammerling, Katzfuss and Smith, 2017](#); [Katzfuss, Hammerling and Smith, 2017](#)) also offers a flexible way to insert different layers of complexity present in climate D&A problems (internal natural variability, numerical model uncertainty, observational errors, sampling uncertainty in space and time, etc.). Despite these advances, the EEA domain remains a fairly untouched territory in terms of multivariate EVT. Even recent papers like [Kew et al. \(2019\)](#); [Luu et al. \(2018\)](#); [Otto et al. \(2018\)](#) and [King \(2017\)](#) are based on univariate EVT only.

Still, in climatological studies, there is no reason to assume that the dependence structure remains identical over space. For example, heavy rainfall in convective prone regions will spatially differ from regions driven by frontal storms. Examples like the one shown in Table 1 below underscore that the practitioner choice of the region under study and the risk measure used

can lead to very different return periods. In addition, the essence of EEA studies is to detect changes between return periods in factual and counterfactual worlds (see, e.g., [Naveau, Hannart and Ribes, 2020](#), for a recent review on the statistical aspects of this problem). If the spatial dependence structure differs between the factual and counterfactual worlds (e.g., a region could be more prone to convective storms in its factual version), it has an influence on the return level. Hence, past studies based on univariate EVT measures may have overlooked some causal signals by not taking into account the underlying multivariate structure.

Our first objective is to investigate how multivariate EVT could be used for EEA. As extreme events in D&A are mostly expressed in terms of threshold exceedances, like $\{X > v\}$ in [Stott, Stone and Allen \(2004\)](#), this naturally leads to the question of how to integrate the multivariate generalized Pareto distribution (GPD) introduced by [Tajvidi \(1996\)](#) into the EEA framework. This distribution has been tailored to represent extremal behaviors when at least one of the components of the vector of interest is large. The probabilistic properties of the multivariate GPD have been well studied by, among others, [Beirlant et al. \(2004\)](#); [Rootzén and Tajvidi \(2006\)](#); [Falk and Guillou \(2008\)](#); [Ferreira and de Haan \(2014\)](#); [Rootzén, Segers and Wadsworth \(2018a\)](#) and [Rootzén, Segers and Wadsworth \(2018b\)](#), while statistical modeling is more recent ([Huser, Davison and Genton, 2016](#); [Kiriliouk et al., 2019](#)).

In most univariate EEA studies (see [Stott et al., 2016](#), and references therein), two types of probability ratios are considered: the Risk Ratio $\frac{p_1}{p_0}$ and the so-called Fraction of Attributable Risk (FAR), defined by

$$\text{FAR} = 1 - \frac{p_0}{p_1},$$

where $p_0 = \mathbb{P}(X > v)$ corresponds to the probability of exceeding the threshold v in the counterfactual world, while p_1 represents the same quantity in the factual world. Using the counterfactual theory of [Pearl \(2000\)](#), the FAR corresponds to the probability of necessary causation, i.e., anthropogenic forcings are necessary for the extreme event to occur, but might not be sufficient. Within the Gaussian set-up, [Hannart et al. \(2016\)](#) and [Hannart and Naveau \(2018\)](#) highlighted the link between causality theory and event attribution studies. The second objective of our work is to explain how Pearl's counterfactual theory can be applied within a multivariate GPD framework, and to identify conditions that maximize the probability of causality, a fundamental feature in any EEA analysis.

The rest of the paper is structured as follows. Section 2 summarizes the relevant background of EEA and the multivariate GPD. Section 3 discusses

the behaviour of univariate probabilities of causation as a function of the threshold v . In Section 4, we make the link between multivariate GPDs and causality by maximizing necessary causation for any linear projection and we discuss the inference strategy. Finally, in Section 5, the proposed methods are applied to weekly winter maxima of precipitation outputs from the French CNRM model that are part of the recent CMIP6 experiment. A discussion and outlook is given in Section 6. Technical details are deferred to the supplementary material (Kiriliouk and Naveau, 2020).

2. Background.

2.1. *Climate event attribution and counterfactual theory.* The question of attribution in EEA is inherently rooted in causality assessment. Pearl (2000) proposed a framework to connect the probabilities p_0 and p_1 to different forms of causality. If E denotes an event (e.g. the 2003 European heatwave) and C its potential cause (e.g. the increase of GHG emissions), three distinct forms of causality are of interest:

1. probability of necessary causation (PN): C is required for E but other factors might be required as well;
2. probability of sufficient causation (PS): C always triggers E but E might occur without C ;
3. probability of necessary and sufficient causation (PNS): both of the above hold.

Hannart et al. (2016) recalled the mathematical definition of these probabilities,

$$\begin{aligned} \text{PN} &= \mathbb{P}[\overline{E} \mid \text{do}(\overline{C}), C, E], & \text{PS} &= \mathbb{P}[E \mid \text{do}(C), \overline{C}, \overline{E}], \\ \text{PNS} &= \mathbb{P}[E \mid \text{do}(C), \overline{E} \mid \text{do}(\overline{C})], \end{aligned}$$

where $\text{do}(C)$ means that the cause is interventionnaly forced to occur. For climate EEA, the cause C can be defined as the presence of anthropogenic forcings. In this setting, Hannart et al. (2016) exploited that E is monotonous with respect to C (the event is more likely when C is present) and C is exogenous (i.e., it is not influenced by any other observed variables). The causation probabilities then simplify to

$$\begin{aligned} (2.1) \quad \text{PN} &= \max \left(1 - \frac{p_0}{p_1}, 0 \right), & \text{PS} &= \max \left(1 - \frac{1 - p_1}{1 - p_0}, 0 \right), \\ \text{PNS} &= \max(p_1 - p_0, 0), \end{aligned}$$

where $p_0 = \mathbb{P}[E \mid \bar{C}]$ corresponds to the probability of E in the counterfactual world and $p_1 = \mathbb{P}[E \mid C]$ to the probability of the same event in the factual world. If $p_0 < p_1$, the PN coincides with the FAR used by [Stott, Stone and Allen \(2004\)](#). In the remainder of this paper, we will use the notation PN (instead of FAR) to highlight its causal interpretation. By construction, one has $\text{PNS} \leq \min(\text{PN}, \text{PS})$ and hence it is worth noticing that a low PNS does not imply the absence of a causal relationship.

A fundamental step in any causality analysis is the definition of the event E . In EEA, the event E is classically defined as the occurrence of some climatological index (e.g. a spatial average over a given region) exceeding a high threshold v ,

$$E = \{\mathbf{w}^T \mathbf{X} > v\},$$

where $\mathbf{X} = (X_1, \dots, X_d)^T$ is a random vector defined on d grid points and $\mathbf{w} = (w_1, \dots, w_d)^T$ are non-negative weights. Let $\mathbf{X}^{(0)}$ and $\mathbf{X}^{(1)}$ denote the climatological vector \mathbf{X} in the counterfactual and the factual world respectively. Hence, the probabilities p_i in (2.1) become

$$p_i = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(i)} > v], \quad i \in \{0, 1\}.$$

Generally, $\mathbf{w}^T \mathbf{X}^{(i)}$ is modelled as a univariate random variable (see, e.g. [Kew et al., 2019](#); [Luu et al., 2018](#); [Otto et al., 2018](#); [King, 2017](#)) and the dependence among the components of $\mathbf{X}^{(i)}$ is ignored. One objective of this paper is to explore how the multivariate dependence among the $\mathbf{X}^{(i)}$'s affects the values p_i , and consequently the causal evidence expressed with PN, PS or PNS, especially if the weights are well chosen. In the next section, we present a simple example to gain some intuition on the difference in return levels between univariate and multivariate modeling.

2.2. Impact of multivariate extremal modeling on return periods. In this section, we illustrate how the dependence structure of \mathbf{X} impacts the return periods of $\mathbf{w}^T \mathbf{X}$. For clarity's sake, we assume that the margins X_i of \mathbf{X} follow unit exponential distributions (in line with the multivariate EVT model of Section 2.3). Hence, the associated return level at each grid point is simply $u_T = \log T$ for a given return period T , i.e., $\mathbb{P}(X_i > u_T) = 1/T$ for $i = 1, \dots, d$. If the X_i 's exhibit complete dependence, i.e., $X_i = X_j$ with probability 1 for all i, j , the return time of $\mathbf{w}^T \mathbf{X}$ is identical to that of X_i , and is simply equal to T . In the opposite case, the margins of \mathbf{X} are independent and by construction, $\mathbf{w}^T \mathbf{X}$ has unit mean and variance $w_1^2 + \dots + w_d^2$. If $w_i = 1/d$ for all i , the return time of $\mathbf{w}^T \mathbf{X}$ associated with u_T corresponds to the quantile of a gamma distribution. The first and third rows in Table 1 show that the return periods of the event $\{w_1 X_1 + w_2 X_2 > u_T\}$ are

much larger than the ones obtained in the complete dependence case. This effect increases as the time period increases and/or the dimension d increases (tables available upon request). Hence, *given the same univariate variable $\mathbf{w}^T \mathbf{X}$ and the same return level $u_T = \log T$, the degree of dependence in the original data greatly influences the return period of the same event.* As the field of EEA is rooted in the computation of events like $\{\mathbf{w}^T \mathbf{X} > u_T\}$ for large T , modeling of multivariate extremal dependence becomes paramount. In the last decades, Rootzén and his colleagues have proposed models and inferential schemes for high return levels that take this dependence into account. The row “Intermediate dependence (MGPD)” in Table 1 concerns such a multivariate model (to be detailed in Section 2.3).

	$w_1 = w_2 = 0.5$		
	$T = 10$	$T = 50$	$T = 100$
Complete dependence	10	50	100
Intermediate dependence (MGPD)	19	96	191
Independence	18	283	979
	$w_1 = 0.2, w_2 = 0.8$		
	$T = 10$	$T = 50$	$T = 100$
Complete dependence	10	50	100
Intermediate dependence (MGPD)	18	88	175
Independence	13	100	237

TABLE 1

Return periods T in years of the event $\{w_1 X_1 + w_2 X_2 > u_T\}$ with $u_T = \log T$ for $T \in \{10, 50, 100\}$ for $w_1 = w_2 = 0.5$ (top) and $w_1 = 0.2, w_2 = 0.8$ (bottom).

The marginal distributions of X_1 and X_2 are unit exponential satisfying $P(X_1 > u_T) = P(X_2 > u_T) = 1/T$. The multivariate GPD has tail dependence coefficient $\chi = 0.5$, see Section 2.3.

Comparing the upper part of Table 1 with the lower part, we see that the choice of the weights also plays a non-negligible role in the return times values. In EEA, one wishes to contrast the factual and counterfactual worlds. Table 1 shows that differences between $p_0 = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(0)} > v]$ and $p_1 = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(1)} > v]$ not only stem from the spatial dependences structures of $\mathbf{X}^{(0)}$ and $\mathbf{X}^{(1)}$, but also from the weight selection.

2.3. The multivariate generalized Pareto distribution. When $\mathbf{X} = X \in \mathbb{R}$, univariate peaks-over-thresholds approaches (Davison and Smith, 1990) consist of choosing a large threshold u and fitting $[X - u \mid X > u]$ to a univariate GPD. Hence, regardless of the underlying distribution of the climatological index X , the GPD can be used to model causation probabilities

of the event $\{X > v\}$ as long as $v > u$. Similarly, a multivariate GPD approximates the tail behavior of $[\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u}]$, where $\mathbf{X} \not\leq \mathbf{u}$ means that at least one component of \mathbf{X} exceeds the corresponding component of the threshold $\mathbf{u} \in \mathbb{R}^d$. In the following, we will see how it can be used to model causation probabilities of the event $\{\mathbf{w}^T \mathbf{X} > v\}$ for $v > \mathbf{w}^T \mathbf{u}$.

From a mathematical point of view, multivariate GPD vectors can be viewed as the limiting solution of any linearly rescaled multivariate vector given that at least one component is large. This asymptotic result can be interpreted as a multivariate version of the Pickands–Balkema–de Haan theorem (Pickands, 1975; Balkema and de Haan, 1974). For the sake of clarity and concision, we will only recall how multivariate GPDs can be simulated and how basic principles are derived from this stochastic definition. The reader interested by theoretical aspects of multivariate GPDs is referred to Rootzén and Tajvidi (2006).

The basic building block to construct standardized multivariate GPD vectors (Rootzén, Segers and Wadsworth, 2018a) is the stochastic representation

$$(2.2) \quad \mathbf{Z}^* \stackrel{\text{d}}{=} E + \mathbf{T} - \max_{1 \leq j \leq d} T_j,$$

where E is a unit exponential random variable and $\mathbf{T} = (T_1, \dots, T_d)^T$ represents any d -dimensional random vector independent of E . One can check that each positive conditional margin has a unit exponential survival function,

$$\mathbb{P}[Z_j^* > z \mid Z_j^* > 0] = \exp(-z), \quad \text{for any } z > 0 \text{ and } j \in \{1, \dots, d\}.$$

Model (2.2) can be generalized by setting, for $\boldsymbol{\sigma} > \mathbf{0}$ and $\boldsymbol{\gamma} \in \mathbb{R}^d$,

$$(2.3) \quad \mathbf{Z} \stackrel{\text{d}}{=} \frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}} (\exp(\boldsymbol{\gamma} \mathbf{Z}^*) - 1),$$

where operations like $\frac{\boldsymbol{\sigma}}{\boldsymbol{\gamma}}$ have to be understood componentwise. We then denote $\mathbf{Z} \sim \text{MGPD}(\mathbf{T}, \boldsymbol{\sigma}, \boldsymbol{\gamma})$. Equation (2.3) implies

$$\mathbb{P}[Z_j > z \mid Z_j > 0] = \overline{H}(z; \sigma_j, \gamma_j), \quad \text{for any } z > 0 \text{ and } j \in \{1, \dots, d\},$$

where $\overline{H}(z; \sigma, \gamma) = (1 + \gamma z / \sigma)_+^{-1/\gamma}$ denotes the survival function of the univariate GPD with scale parameter $\sigma > 0$ and shape parameter $\gamma \in \mathbb{R}$. Hence, the conditional margins $[Z_j \mid Z_j > 0]$ follow univariate GPDs¹.

The random “generator” \mathbf{T} in (2.2) drives the extremal dependence of \mathbf{Z} , often summarized by the *tail dependence coefficient* $\chi \in [0, 1]$ (for $d = 2$). If

¹although the random variables Z_1, \dots, Z_d may not follow GPDs themselves.

F_1, F_2 denote the unconditional marginal distribution functions of Z_1, Z_2 , then χ measures the probability of $F_1(Z_1)$ being large given that $F_2(Z_2)$ is large as the threshold increases,

$$\begin{aligned}\chi &= \lim_{q \uparrow 1} \mathbb{P}[F_1(Z_1) > q \mid F_2(Z_2) > q] \\ &= \mathbb{E} \left[\min \left(\frac{e^{T_1 - \max(T_1, T_2)}}{\mathbb{E}[e^{T_1 - \max(T_1, T_2)}]}, \frac{e^{T_2 - \max(T_1, T_2)}}{\mathbb{E}[e^{T_2 - \max(T_1, T_2)}]} \right) \right].\end{aligned}$$

A large value of χ corresponds to strong extremal dependence between Z_1 and Z_2 , whereas $\chi = 0$ corresponds to tail independence. For more details on χ in the context of multivariate GPDs, see the supplementary material in [Kiriliouk et al. \(2019\)](#). As an example, Figure 1 displays 500 bivariate random draws from a multivariate GPD model where \mathbf{T} is zero-mean bivariate Gaussian with unit covariance matrix I_2 , corresponding to $\chi = 0.6$.

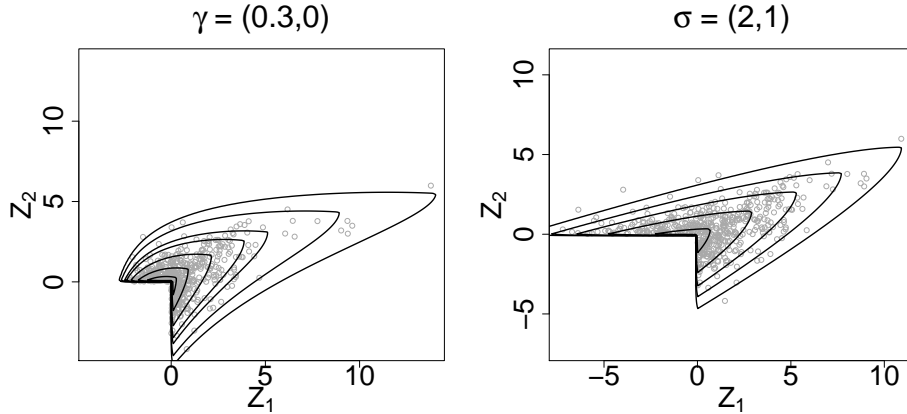


Fig 1: Scatterplots and density contours from 500 bivariate GPD random draws using (2.2) and (2.3) with parameters $\gamma = (0.3, 0)$, $\sigma = (1, 1)$ for the left panel and $\gamma = (0, 0)$, $\sigma = (2, 1)$ for the right panel. The generator \mathbf{T} is zero-mean bivariate Gaussian with unit covariance matrix I_2 .

To make the link with the probabilities p_0 and p_1 used for EEA, we need a tool to project the information contained in a possibly complex multivariate GPD structure into a single valued summary. The following proposition provides this key tool.

PROPOSITION 2.1 (Linear-projection, Rootzén, Segers and Wadsworth (2018a)). *If $\mathbf{Z} \sim \text{MGPD}(\mathbf{T}, \boldsymbol{\sigma}, \gamma)$ with $\gamma = \gamma \mathbf{1}_d$, then for any non-negative weights $\mathbf{w} = (w_1, \dots, w_d)^T$ such that $\mathbb{P}[\mathbf{w}^T \mathbf{Z} > 0] > 0$, the linear projection of \mathbf{Z} , conditioned on being positive, follows a univariate GPD, i.e.,*

$$[\mathbf{w}^T \mathbf{Z} \mid \mathbf{w}^T \mathbf{Z} > 0] \sim \text{GPD}(\mathbf{w}^T \boldsymbol{\sigma}, \gamma).$$

As explained in Section 2.1, our main inferential objective will be to estimate the probability $\mathbb{P}[\mathbf{w}^T \mathbf{X} > v]$. A priori, if \mathbf{X} represents a climatological index on d grid points, it does not follow a multivariate GPD, but, since $\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u} \approx \mathbf{Z}$, we will use Proposition 2.1 to deduce a suitable model for \mathbf{X} in Section 4.

3. Causation probabilities for univariate extremes. To understand first how PN, PS and PNS behave for univariate extremes, we take $p_0(v) = \mathbb{P}[X^{(0)} > v]$ and $p_1(v) = \mathbb{P}[X^{(1)} > v]$ when $X^{(0)}$ and $X^{(1)}$ are either Gaussian or GPD random variables. The left panel of Figure 2 shows the case where the counterfactual world corresponds to a standardized Gaussian variable, $X^{(0)} \sim N(0, 1)$, and the factual world is one Kelvin warmer with a higher variability, $X^{(1)} \sim N(1, 1.5)$. This artificial design mimics the typical behaviour of mean temperature anomalies. Two features can be highlighted from this example: PN goes to one as v increases, and the maximum of PNS is around 0.35. In other words, the probability of necessary causation becomes certain for extremes (large v), and the probability of necessary and sufficient causation can be reasonably high in the Gaussian case.

To contrast these remarks with other types of tail behaviors, the middle panel of Figure 2 displays a GPD case with equal shape parameter $\gamma = 0.2$ in the counterfactual and factual worlds, but different scale parameters, $\sigma^{(0)} = 1$ and $\sigma^{(1)} = 1.5$. One can see that, as v increases, PN converges to a constant around $0.7 < 1$, and PNS remains small for any value of v . Hence, causal evidencing is much more difficult than in the Gaussian set-up, where a rare event in the factual world (p_1 small) would be nearly impossible in the counterfactual world (p_0 almost zero). In contrast, even a very rare event in the factual world will not be impossible in a GPD counterfactual world. Concerning PNS, it is small in the second panel and this phenomenon is even more pronounced when the shape parameter changes between the two worlds; see the right panel where $\gamma^{(0)} = 0.1$, $\gamma^{(1)} = 0.3$, and $\sigma^{(0)} = \sigma^{(1)} = 1$. As PNS is near zero for all v , there is no reason to maximize it. Instead, maximizing causality will correspond to maximizing PN in the remainder of this work.

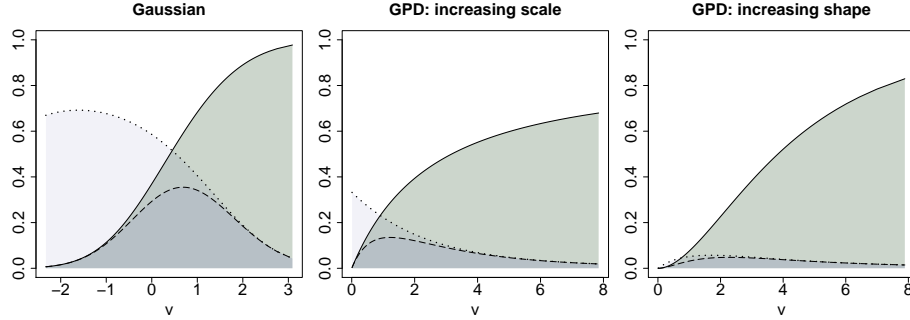


Fig 2: Probabilities of necessary causation (PN, solid line), sufficient causation (PS, dotted line) and necessary and sufficient causation (PNS, dashed line) as functions of v . Left panel: Gaussian set-up, $N(0, 1)$ for the counterfactual world and $N(1, 1.5)$ for the factual one. Middle and right panels: GPD set-up, $\text{GPD}(1, 0.2)$ for the counterfactual world, $\text{GPD}(1.5, 0.2)$ for the factual one (middle) and $\text{GPD}(1, 0.1)$ for the counterfactual world, $\text{GPD}(1, 0.3)$ for the factual one (right).

In practice, $X^{(0)}$ and $X^{(1)}$ do not follow GPDs. Using a classical peaks-over-thresholds approach, we can condition on some high threshold $u^{(i)}$ and approximate the probabilities $p_i(v) = \mathbb{P}[X^{(i)} > v]$ for $v > u^{(i)}$ by

$$(3.1) \quad p_i(v) \approx \mathbb{P}[X^{(i)} > u^{(i)}] \overline{H}\left(v - u^{(i)}; \sigma^{(i)}, \gamma^{(i)}\right), \quad \text{for } i \in \{0, 1\}.$$

We can now formalize the tail behaviour observed in Figure 2. Whenever the limit of $\text{PN}(v)$ for large v is finite², it has to be equal to

$$(3.2) \quad \begin{cases} \mathbb{1}\left\{\gamma^{(0)} < \gamma^{(1)}\right\} & \text{if } \gamma^{(1)} \neq \gamma^{(0)}, \\ 1 - \frac{p_0(u^{(0)})}{p_1(u^{(1)})} \left(\frac{\sigma^{(0)}}{\sigma^{(1)}}\right)^{1/\gamma} & \text{if } \gamma^{(1)} = \gamma^{(0)} =: \gamma, \gamma \neq 0, \\ \mathbb{1}\left\{\sigma^{(0)} < \sigma^{(1)}\right\} & \text{if } \gamma^{(1)} = \gamma^{(0)} = 0. \end{cases}$$

where $\mathbb{1}(A)$ represents the indicator function, equal to one if A is true and zero otherwise.

²Degenerate cases can occur when $p_1(v) = 0$. For example, if $\gamma^{(1)} < 0$, the PN is not defined for $v \geq u^{(1)} - \sigma^{(1)}/\gamma^{(1)}$, which is visible for the dashed line in the left panel of Figure 3.

The left panel of Figure 3 shows how three different GPD shape parameters, $\gamma = -0.4$, $\gamma = 0$ and $\gamma = 0.4$ (dashed, solid and dotted lines respectively) influence the increase of the PN with respect to the threshold v . The right panel of Figure 3 points out possible atypical behaviours of the PN, highlighting that $\text{PN}(v)$ is not always increasing as v increases. Here, the solid line corresponds to a counterfactual world with $(\gamma^{(0)}, \sigma^{(0)}) = (0, 2)$ compared to a factual world with $(\gamma^{(1)}, \sigma^{(1)}) = (0.4, 1)$, while the dotted line represents the converse change: from $(\gamma^{(0)}, \sigma^{(0)}) = (0.4, 1)$ to $(\gamma^{(1)}, \sigma^{(1)}) = (0, 2)$.

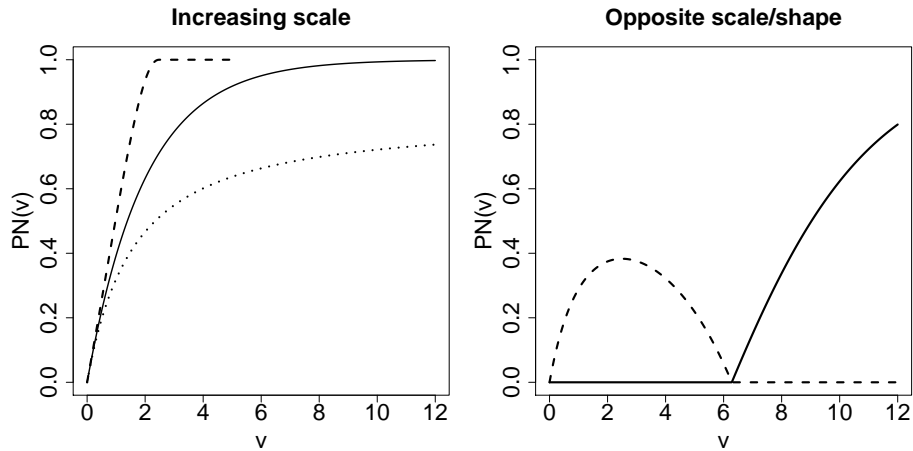


Fig 3: Probability of necessary causation as a function of v , see (2.1) and (3.1). Left panel: the counterfactual scale is $\sigma^{(0)} = 1$, increasing to $\sigma^{(1)} = 2$ in the factual world, while the shape parameter is identical, $\gamma^{(0)} = \gamma^{(1)}$, equal to -0.4 , 0 , and 0.4 for the dashed, solid and dotted lines respectively. Right panel: the solid line corresponds to increasing shape and decreasing scale, $(\gamma^{(0)}, \sigma^{(0)}) = (0, 2)$ and $(\gamma^{(1)}, \sigma^{(1)}) = (0.4, 1)$. The dashed line corresponds to the opposite scenario : $(\gamma^{(0)}, \sigma^{(0)}) = (0.4, 1)$ and $(\gamma^{(1)}, \sigma^{(1)}) = (0, 2)$.

4. Necessary causation in a multivariate set-up.

4.1. *The multivariate GPD and necessary causation.* Let \mathbf{X} be any d -dimensional random vector such that $[\mathbf{X} - \mathbf{u} \mid \mathbf{X} \not\leq \mathbf{u}] \approx \mathbf{Z}$ for some high threshold $\mathbf{u} \in \mathbb{R}^d$, where $\mathbf{Z} \sim \text{MGPD}(\mathbf{T}, \boldsymbol{\sigma}, \boldsymbol{\gamma})$. Then, according to Proposition 2.1, the extremal information contained in any linear projection $\mathbf{w}^T \mathbf{X}$ can be approximated, up to a normalizing constant, by a univariate GPD

survival function. More precisely, for any $v > \mathbf{w}^T \mathbf{u}$, we can write

$$(4.1) \quad \begin{aligned} \mathbb{P}[\mathbf{w}^T \mathbf{X} > v] &= \mathbb{P}[\mathbf{w}^T \mathbf{X} > \mathbf{w}^T \mathbf{u}] \mathbb{P}[\mathbf{w}^T (\mathbf{X} - \mathbf{u}) > v - \mathbf{w}^T \mathbf{u} \mid \mathbf{w}^T (\mathbf{X} - \mathbf{u}) > 0] \\ &\approx \mathbb{P}[\mathbf{w}^T \mathbf{X} > \mathbf{w}^T \mathbf{u}] \bar{H}(v - \mathbf{w}^T \mathbf{u}; \mathbf{w}^T \boldsymbol{\sigma}, \gamma), \end{aligned}$$

for $w_1 + \dots + w_d = 1$ and $\boldsymbol{\gamma} = \gamma \mathbf{1}_d$. Constraining the weights to sum to a constant is necessary to ensure identification of $\boldsymbol{\sigma}$. The condition $\boldsymbol{\gamma} = \gamma \mathbf{1}_d$ implies that conditional marginal distributions $[Z_j \mid Z_j > 0]$ have equal shape parameters for $j = 1, \dots, d$. Therefore, homogeneous spatial regions (in terms of the shape parameter) have to be identified in practice. This is closely related to the regional frequency analysis problem treated in hydrology (Carreau, Naveau and Neppel, 2017). Finally, we note that the dependence structure of \mathbf{X} is present in the term $\mathbb{P}[\mathbf{w}^T \mathbf{X} > \mathbf{w}^T \mathbf{u}]$ only.

Any linear projection in the factual and counterfactuals worlds, denoted $p_1(v; \mathbf{w}) = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(1)} > v]$ and $p_0(v; \mathbf{w}) = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(0)} > v]$ respectively, can now be used to compute a probability of necessary causation that depends on the weight \mathbf{w} and the dependence structure of $\mathbf{X}^{(0)}$ and $\mathbf{X}^{(1)}$,

$$(4.2) \quad \text{PN}(v, \mathbf{w}) = \max \left(1 - \frac{\mathbb{P}[\mathbf{w}^T \mathbf{X}^{(0)} > v]}{\mathbb{P}[\mathbf{w}^T \mathbf{X}^{(1)} > v]}, 0 \right),$$

where $\mathbf{X}^{(i)}$ satisfies approximation (4.1) for $i \in \{0, 1\}$. To understand how dependence affects the strength of necessary causation, we study the value of $\text{PN}(v, \mathbf{w})$ in the bivariate case with $\mathbf{w} = (0.5, 0.5)$, $\mathbf{X}^{(0)} \stackrel{d}{=} \mathbf{Z}^{(0)} \sim \text{MGPD}(\mathbf{T}^{(0)}, \boldsymbol{\sigma}^{(0)}, \mathbf{0})$ and $\mathbf{X}^{(1)} \stackrel{d}{=} \mathbf{Z}^{(1)} \sim \text{MGPD}(\mathbf{T}^{(1)}, \boldsymbol{\sigma}^{(1)}, \mathbf{0})$. In the example displayed in Figure 4, two different dependence structures are investigated, summarized by the tail dependence coefficient χ . The dotted line corresponds to increasing dependence, from $\chi^{(0)} = 0.3$ to $\chi^{(1)} = 0.5$ and an increasing marginal scale, from $\boldsymbol{\sigma}^{(0)} = (1, 1)$ to $\boldsymbol{\sigma}^{(1)} = (2, 2)$. The dashed line again represents increasing dependence, but decreasing marginal scale, from $\boldsymbol{\sigma}^{(0)} = (2, 2)$ to $\boldsymbol{\sigma}^{(1)} = (1, 1)$. Finally, the solid line shows increasing marginal scale of the same order as the dotted line, and decreasing dependence, from $\chi^{(0)} = 0.5$ to $\chi^{(1)} = 0.3$. Figure 4 shows that the dependence structure can have an impact on the PN for any finite value of v . In other words, EEA based on a hypothesis of independence (eg, in space) will lead to incomplete statements concerning the strength of PN whenever the multivariate extremes are dependent. Figure 4 also suggests that, as v increases, the impact of an increasing dependence in the factual world becomes negligible. However, it is important to keep in mind that in applications, the marginal scale might be constant between the two worlds. In that case, we will see the impact of increasing dependence in Figure 6.

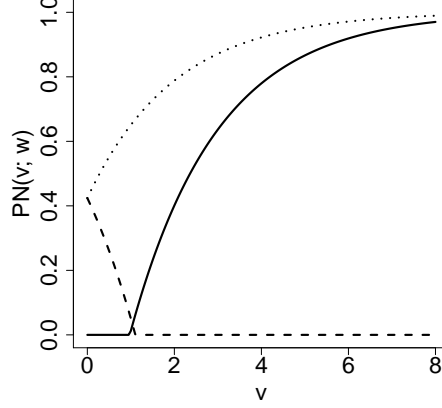


Fig 4: $\text{PN}(v, \mathbf{w} = (0.5, 0.5))$ defined by (4.2) between two bivariate GPDs, $\mathbf{Z}^{(0)} \sim \text{MGPD}(\mathbf{T}^{(0)}, \boldsymbol{\sigma}^{(0)}, \mathbf{0})$ and $\mathbf{Z}^{(1)} \sim \text{MGPD}(\mathbf{T}^{(1)}, \boldsymbol{\sigma}^{(1)}, \mathbf{0})$. The dotted line corresponds to $\chi^{(0)} = 0.3$, $\chi^{(1)} = 0.5$, $\boldsymbol{\sigma}^{(0)} = (1, 1)$ and $\boldsymbol{\sigma}^{(1)} = (2, 2)$. The dashed line differs from the dotted line by $\boldsymbol{\sigma}^{(0)} = (2, 2)$ and $\boldsymbol{\sigma}^{(1)} = (1, 1)$. The solid line differs from the dotted line by $\chi^{(0)} = 0.5$ and $\chi^{(1)} = 0.3$.

4.2. *Maximizing necessary causation.* In a multivariate Gaussian set-up, Hannart and Naveau (2018) proposed to maximize causation probabilities by using the linear projection that will contrast the factual and counterfactual worlds the most. Their solution was similar to linear discriminant analysis. This leads to the question of how to reduce the dimension of a multivariate GPD vector, while ensuring that the projected data contains the most information in terms of causality for extremes.

More specifically, the choice of \mathbf{w} plays an essential role in the maximization of necessary causation for multivariate GPD random variables. To address this point in the bivariate case, we need the following result.

PROPOSITION 4.1. *Let $\gamma \in \mathbb{R}$ and consider two positive bivariate scale parameters: $\boldsymbol{\sigma}^{(0)} = (\sigma_1^{(0)}, \sigma_2^{(0)})^T$ and $\boldsymbol{\sigma}^{(1)} = (\sigma_1^{(1)}, \sigma_2^{(1)})^T$. Denote*

$$R = \frac{\left(\sigma_1^{(0)} \sigma_2^{(1)} - \sigma_2^{(0)} \sigma_1^{(1)} \right) \left(\sigma_1^{(0)} \sigma_2^{(1)} - \sigma_2^{(0)} \sigma_1^{(1)} + \gamma v \left\{ (\sigma_2^{(1)} - \sigma_1^{(1)}) - (\sigma_2^{(0)} - \sigma_1^{(0)}) \right\} \right)}{(\sigma_2^{(1)} - \sigma_1^{(1)})(\sigma_2^{(0)} - \sigma_1^{(0)}) \left\{ (\sigma_2^{(1)} - \sigma_1^{(1)}) - (\sigma_2^{(0)} - \sigma_1^{(0)}) \right\}^2},$$

and if $R > 0$, define the weights

$$w_{\pm}(v) = \frac{\sigma_2^{(1)} - \sigma_2^{(0)}}{(\sigma_2^{(1)} - \sigma_1^{(1)}) - (\sigma_2^{(0)} - \sigma_1^{(0)})} \pm \sqrt{R}.$$

If $R \geq 0$ and if one of the two weights $w_{\pm}(v)$ belongs to $(0, 1)$, then this weight, denoted w_{opt} , maximizes

$$(4.3) \quad \left\{ 1 - \frac{\overline{H}\left(v; w\sigma_1^{(0)} + (1-w)\sigma_2^{(0)}, \gamma\right)}{\overline{H}\left(v; w\sigma_1^{(1)} + (1-w)\sigma_2^{(1)}, \gamma\right)} \right\}.$$

In all other cases, only zero or unit weights maximize this ratio.

When $\gamma = 0$, $w_{\pm}(v)$ is simpler because it does not depend on v . Expression (4.3) is an approximation of the PN defined in (4.2); it is equal to the PN when $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}$ are multivariate GPDs and when $\mathbb{P}[\mathbf{w}^T \mathbf{Z}^{(0)} > 0] = \mathbb{P}[\mathbf{w}^T \mathbf{Z}^{(1)} > 0]$, i.e., when the dependence structure remains constant between the two worlds. Proposition 4.1 allows us to study the gain in terms of PN with respect to the weight w . When unit or zero weights are chosen as the optimal solution in Proposition 4.1, only one coordinate is considered and no linear projection is necessary. This happens when the contrast in one of the margins between the factual and counterfactual world is already sufficient to optimize PN. However, Proposition 4.1 shows that, to maximize necessary causality, one needs to consider only those components that (individually) give the largest PN. As an example, take $\mathbf{X}^{(0)} \stackrel{d}{=} \mathbf{Z}^{(0)} \sim \text{MGPD}(\mathbf{T}^{(0)}, (1, 2)^T, \gamma \mathbf{1}_d)$ and $\mathbf{X}^{(1)} \stackrel{d}{=} \mathbf{Z}^{(1)} \sim \text{MGPD}(\mathbf{T}^{(1)}, (1.5, 2)^T, \gamma \mathbf{1}_d)$. The dependence structures of $\mathbf{T}^{(1)}$ and $\mathbf{T}^{(0)}$ are chosen such that $\chi^{(0)} = \chi^{(1)} = 0.5$. Hence, the difference between the two worlds is only due to the scale change in one of the components. Figure 5 shows the PN gain as a function of v , i.e., the ratio between $\text{PN}(v, (0.5, 0.5)^T)$ and $\text{PN}(v, (w_{\text{opt}}, 1 - w_{\text{opt}})^T)$ where $w_{\text{opt}} = 1$ based on Proposition 4.1. Each curve corresponds to a different shape parameter (with constraint $\gamma^{(0)} = \gamma^{(1)}$), equal to -0.4 (dashed line), 0 (solid line), and 0.4 (dotted line), respectively. We see that the optimal weight can lead to a large increase in necessary causality, particularly when the shape parameter is positive (dotted line).

Explicit optimal weights like in Proposition 4.1 can only be obtained in very specific cases. For example, for the bivariate Gaussian GPD with $\gamma = 1$, the probability $\mathbb{P}[\mathbf{w}^T \mathbf{Z}^{(i)} > 0]$ does not depend on \mathbf{w} (see Proposition A.1 in the supplementary material, Kiriliouk and Naveau (2020)). For most other

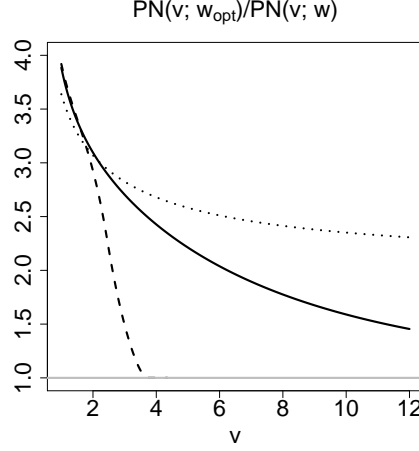


Fig 5: Necessary causation gain for $\mathbf{X}^{(0)} \stackrel{d}{=} \mathbf{Z}^{(0)} \sim \text{MGPD}(\mathbf{T}^{(0)}, (1, 2)^T, \gamma \mathbf{1}_d)$ and $\mathbf{X}^{(1)} \stackrel{d}{=} \mathbf{Z}^{(1)} \sim \text{MGPD}(\mathbf{T}^{(1)}, (1.5, 2)^T, \gamma \mathbf{1}_d)$, where $\mathbf{T}^{(1)}, \mathbf{T}^{(0)}$ are Gaussian random vectors such that $\chi^{(0)} = \chi^{(1)} = 0.5$. The ratio of $\text{PN}(v, (w_{\text{opt}}, 1 - w_{\text{opt}})^T)$ to $\text{PN}(v, (0.5, 0.5)^T)$ is shown as a function of v , where $w_{\text{opt}} = 1$ based on Proposition 4.1. The dashed, solid and dotted lines correspond to a shape parameter (with constraint $\gamma^{(0)} = \gamma^{(1)}$) of -0.4 , 0 and 0.4 respectively.

cases, numerical optimization schemes have to be used, especially beyond the bivariate set-up. In order to move closer to practical applications, we need to couple this optimization procedure with inference in a multivariate context.

4.3. Inference. Let $\mathbf{X}_1^{(0)}, \dots, \mathbf{X}_n^{(0)}$ and $\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_n^{(1)}$ denote two independent samples of size n , representing climate model output in the counterfactual and the factual world respectively, and let $\mathbf{u}^{(0)}, \mathbf{u}^{(1)}$ denote two high thresholds. For $i \in \{0, 1\}$, let N_i denote the number of observations among $\mathbf{X}_1^{(i)}, \dots, \mathbf{X}_n^{(i)}$ that have at least one component exceeding $\mathbf{u}^{(i)}$. Extracting these observations and subtracting $\mathbf{u}^{(i)}$, we obtain the multivariate GPD samples $\mathbf{Z}_1^{(i)}, \dots, \mathbf{Z}_{N_i}^{(i)}$. For $v > \mathbf{w}^T \mathbf{u}^{(i)}$, an estimator of $p_i(v; \mathbf{w}) = \mathbb{P}[\mathbf{w}^T \mathbf{X}^{(i)} > v]$ and hence of the PN follows from approximation (4.1). The first term, $\mathbb{P}[\mathbf{w}^T \mathbf{X}^{(i)} > \mathbf{w}^T \mathbf{u}^{(i)}]$, can be estimated nonparametrically by

$$\hat{p}_i^{(\text{emp})}(v; \mathbf{w}) = \frac{1}{n} \sum_{t=1}^n \mathbb{1} \left\{ \mathbf{w}^T \mathbf{X}_t^{(i)} > v \right\}.$$

To estimate the second term, $\overline{H}(v - \mathbf{w}^T \mathbf{u}^{(i)}; \mathbf{w}^T \boldsymbol{\sigma}^{(i)}, \gamma^{(i)})$, we first compute estimators $\hat{\boldsymbol{\sigma}}^{(i)}$ and $\hat{\gamma}^{(i)}$ by applying the method of probability weighted moments to $(Z_{tj}^{(i)} \mid Z_{tj}^{(i)} > 0)_{t=1, \dots, N_i}$ for $j \in \{1, \dots, d\}$ (see Section C in the supplementary material, Kiriliouk and Naveau (2020)). Next, we set $\hat{\gamma}^{(i)} = d^{-1} \sum_{j=1}^d \hat{\gamma}_j^{(i)}$ ³. Finally, we estimate $p_i(v; \mathbf{w})$ by

$$(4.4) \quad \hat{p}_i(v; \mathbf{w}) = \begin{cases} \hat{p}_i^{(\text{emp})}(v; \mathbf{w}) & \text{if } v \leq \mathbf{w}^T \mathbf{u}^{(i)}, \\ \hat{p}_i^{(\text{emp})}(\mathbf{w}^T \mathbf{u}^{(i)}; \mathbf{w}) \overline{H}(v - \mathbf{w}^T \mathbf{u}^{(i)}; \mathbf{w}^T \hat{\boldsymbol{\sigma}}^{(i)}, \hat{\gamma}^{(i)}) & \text{if } v \leq \mathbf{w}^T \mathbf{u}^{(i)}. \end{cases}$$

Alternatively, we could directly estimate $\gamma^{(i)}$ and $\mathbf{w}^T \boldsymbol{\sigma}^{(i)}$ by applying the method of probability weighted moments to $(\mathbf{w}^T \mathbf{Z}_t^{(i)} \mid \mathbf{w}^T \mathbf{Z}_t^{(i)} > 0)_{t=1, \dots, N_i}$, which reduces uncertainty and enforces the constraint of equal shape parameters, but requires the weights \mathbf{w} to be chosen upfront. Section D in the supplementary material (Kiriliouk and Naveau, 2020) shows a small simulation experiment, confirming the good performance of $\widehat{\text{PN}} = 1 - \hat{p}_0/\hat{p}_1$.

In the previous sections, we studied the increase in PN for changing dependence structures and marginal parameters. Another important question is what happens when marginal parameters do not change ($\boldsymbol{\sigma}^{(0)} = \boldsymbol{\sigma}^{(1)}$ and $\gamma^{(0)} = \gamma^{(1)}$), while dependence increases ($\chi^{(1)} > \chi^{(0)}$). Under a hypothesis of independence in space, one would aggregate the observations from all grid points (i.e., calculate $\mathbf{w}^T \mathbf{X}^{(i)}$) and estimate the univariate PN, thus possibly underestimating the true PN. To see by how much, and how the result varies with the dimension, we conduct the following experiment. Consider $d = 9$ points on a regular 3×3 unit distance grid. For distances from 1 to $\sqrt{8}$, pairwise tail dependence coefficients ranging from 0.4 to 0.3 for the counterfactual world and from 0.55 to 0.4 for the factual world were obtained using a Whittle-Matérn correlation function⁴. We evaluate the PN in the 99% quantile of $\mathbf{w}^T \mathbf{Z}^{(0)}$ using equal weights, calculated based on a pre-simulation run of sample size 10^6 and held fixed. Figure 6 shows boxplots of the multivariate estimates PN minus the univariate estimates, based on 1000 samples of size $n = 2000$. The black line corresponds to the true values, calculated using the formulas in Sections A and B in the supplementary material (Kiriliouk and Naveau, 2020). We see that as the dimension increases, taking dependence into account increases necessary causation.

³An alternative method to enforce equal shape parameters is described in Carreau, Naveau and Neppel (2017)

⁴The covariance matrices $\Sigma^{(0)}$ and $\Sigma^{(1)}$ are generated using a Whittle-Matérn correla-

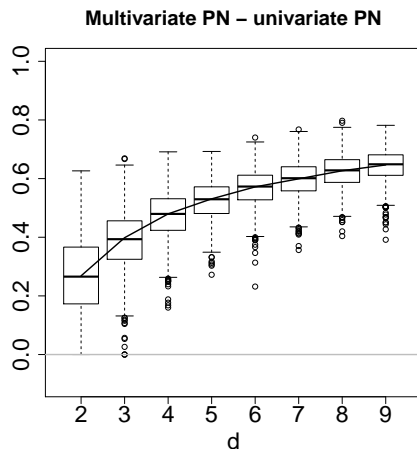


Fig 6: Boxplots of the multivariate estimates $\widehat{\text{PN}} = 1 - \widehat{p}_0/\widehat{p}_1$ minus the univariate PN estimates of aggregated data, where \widehat{p}_i is defined in (4.4), and $d \in \{2, \dots, 9\}$. 1000 samples of size $n = 2000$ were simulated from a multivariate Gaussian GPD model with $\sigma^{(0)} = \sigma^{(1)} = 1$ and $\gamma^{(0)} = \gamma^{(1)} = 0$, $\chi^{(0)} \in [0.3, 0.4]$ and $\chi^{(1)} \in [0.4, 0.55]$ (pairwise). The black line corresponds to the true values.

5. Analysis of heavy precipitation from the CNRM model. Evidencing causality is more difficult for heavy rainfall than for extreme temperatures, because precipitation variability is greater in space and time and because extreme rainfall has heavier tails than temperatures (extreme rainfall often has $\gamma \approx 0.2$, see, e.g. [Katz, Parlange and Naveau \(2002\)](#)). We work with simulated rainfall time series from the French global climate model of Météo-France (CNRM) that belongs to the latest Coupled Model Intercomparison Project (CMIP6). We consider the winter months between the 1st of January 1985 until the 31st of August 2014 over the region defined by -10 to 40 in longitude and 35 to 60 in latitude (corresponding to central Europe). Our factual and counterfactual worlds correspond to two historical runs, the second one of which has only natural forcings. We take the weekly maxima of winter precipitation. As the number of years covers only three decades, the rainfall series can be considered stationary in time within each world. Concerning their spatial structure, we apply the partitioning around medoids (PAM) algorithm ([Kaufman and Rousseeuw, 1990](#)) to the

tion function with fixed shape $\kappa^{(0)} = \kappa^{(1)} = 1$ and varying scales $\phi^{(0)} = 1$, $\phi^{(1)} = 2.5$. The correlation matrices are then multiplied by 10 to obtain $\Sigma^{(0)}$ and $\Sigma^{(1)}$

counterfactual rainfall run. The difference with the original PAM version is that our “distance” between two locations s and t is tailored to threshold exceedances via

$$\hat{d}_{st} = \frac{1 - \hat{\chi}_{st}}{2(3 - \hat{\chi}_{st})},$$

where $\hat{\chi}_{st}$ denotes the standard empirical estimator of the pairwise tail dependence coefficient (see, e.g. Kiriliouk, Segers and Warchol, 2016, and references therein). Our approach is close to the one of Bernard et al. (2013), who focused on maxima instead of threshold exceedances. Figure 7 displays the spatial structure for $K = 40$ clusters⁵. Although no spatial coordinates were given to the algorithm, the clusters appear to be spatially homogeneous and climatologically coherent. As the multivariate GPD is tailored for asymptotic dependence, identifying dependent regions helps to improve its fit. In addition, such a spatial clustering makes the assumption of a constant shape parameter within a region more reasonable. We hence model

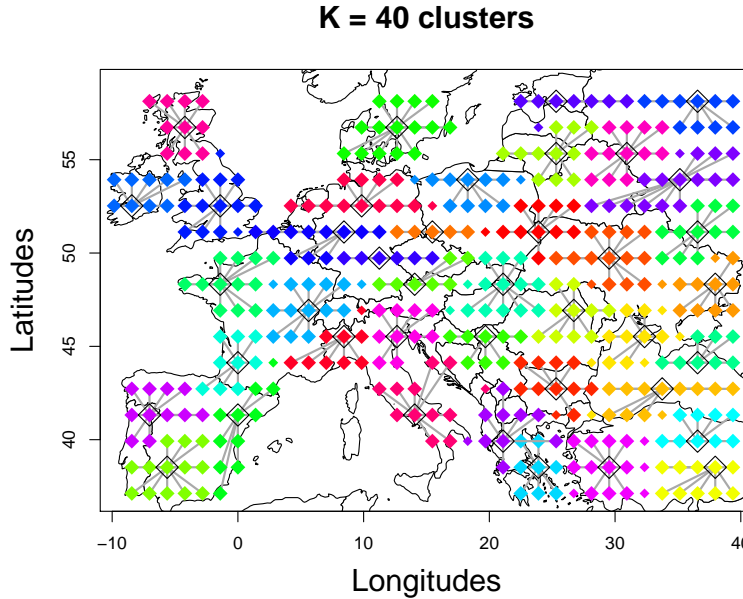


Fig 7: Clustering of weekly maximum winter precipitation in central Europe between January 1985 and August 2014, using the PAM algorithm with distance based on pairwise tail dependence coefficients.

each cluster independently, calculating the estimated multivariate PN based

⁵Other values of K were tested and provided similar patterns.

on \hat{p}_0, \hat{p}_1 defined in (4.4).

Figures 8 and 9 show necessity causation probabilities for the five-year and fifty-year return level respectively. The return levels were calculated based on quantiles of $\mathbf{w}^T X^{(0)}$ for each cluster, with equal weights. Both figures show the PN per cluster, calculated using equal weights (top) and optimal weights (bottom). The diameters of the black circles around the PN estimates are proportional to the length of bootstrap-based 95% confidence intervals. Higher PN does not necessarily come with higher uncertainty; see for instance the cluster around northern Italy, whose confidence interval is narrow for the five-year return level and even more so for the fifty-year return level. Comparing the two panels of Figure 8 shows that the differences between the factual and the counterfactual world are higher when using optimal weights. This feature is even more striking for the fifty-year return levels, see Figure 9. Except for locations near the English channel, most points have a probability of necessary causation that is greater than 0.5. In particular, a few points like northern Italy shows a probability near one.

In Section E of the supplementary material (Kiriliouk and Naveau, 2020), the above approach is compared to several univariate approaches for the five-year return level. We found that, even though patterns are similar (i.e., high PN around northern Italy), a univariate analysis leads to lower PN on average and to wider confidence intervals when PN is relatively high (> 0.5). Hence, a multivariate GPD approach enhances the causality message of a univariate analysis and aids in decreasing the uncertainty of the estimates.

Finally, our analysis is not sufficient to conclude general climatological results about heavy rainfall. The patterns found here may be due to this specific climate model, internal climate variability or other sources of variability. An exhaustive analysis of all the CMIP6 models, in terms of computer resources and climatological expertise, is beyond the scope of this work.

6. Discussion. This paper illustrates that methods combining multivariate extreme-value theory and counterfactual theory could help climatologists working on causality and multivariate extremes (see, e.g. Kim et al., 2016; Zscheischler et al., 2018). An advantage of our approach is its simplicity: we consider an event to be extreme if the weighted average of a climatological random vector exceeds a high threshold, and EVT naturally suggests the multivariate GPD to model such multivariate threshold exceedances. While multivariate EVT models can take on complex parametric forms, the model we propose is easily estimated since linear projections of multivariate GPD vectors behave like a univariate GPD. When spatial dependence changes between the factual and the counterfactual world, a univariate analysis

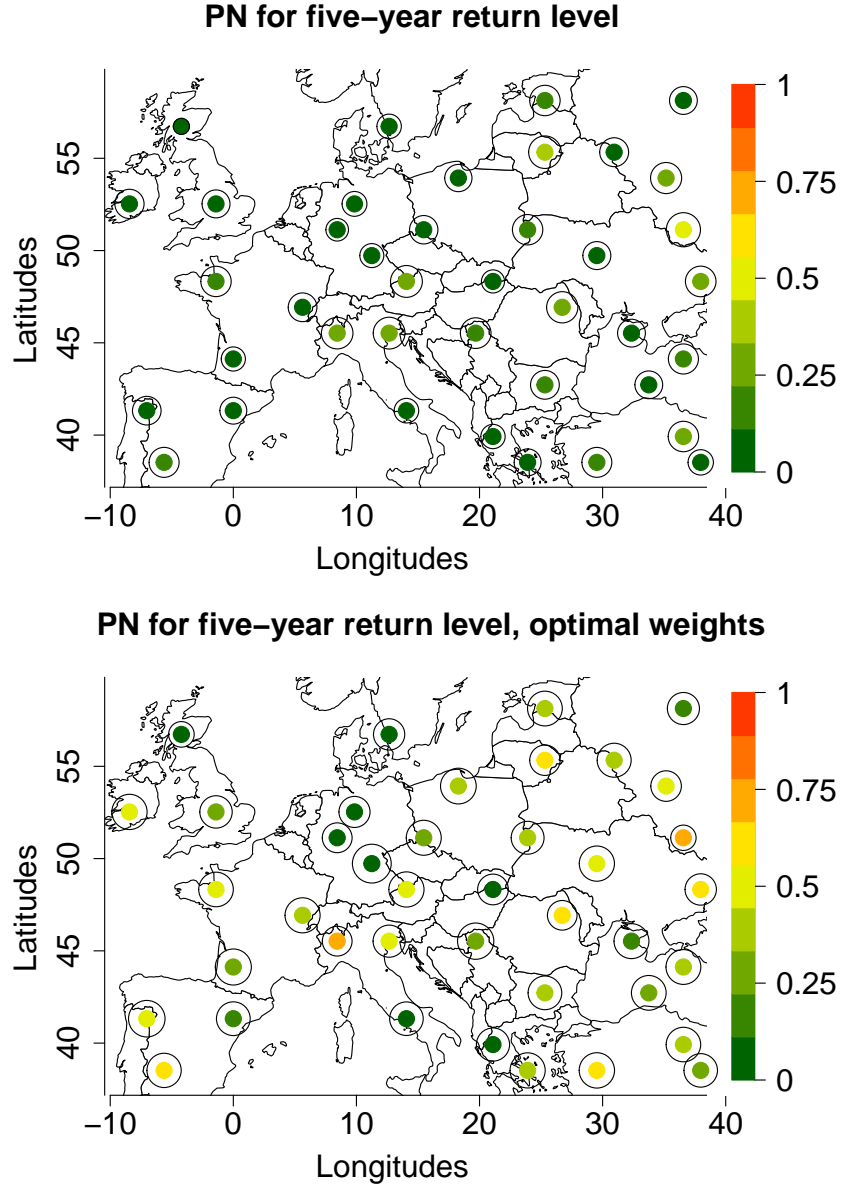


Fig 8: Necessary causation probabilities for a five-year return level of weighted maximal weekly winter precipitation in the counterfactual world for each cluster, calculated using equal weights (top) and optimal weights (bottom). The diameters of the black circles around the estimates are proportional to the length of bootstrap-based 95% confidence intervals.

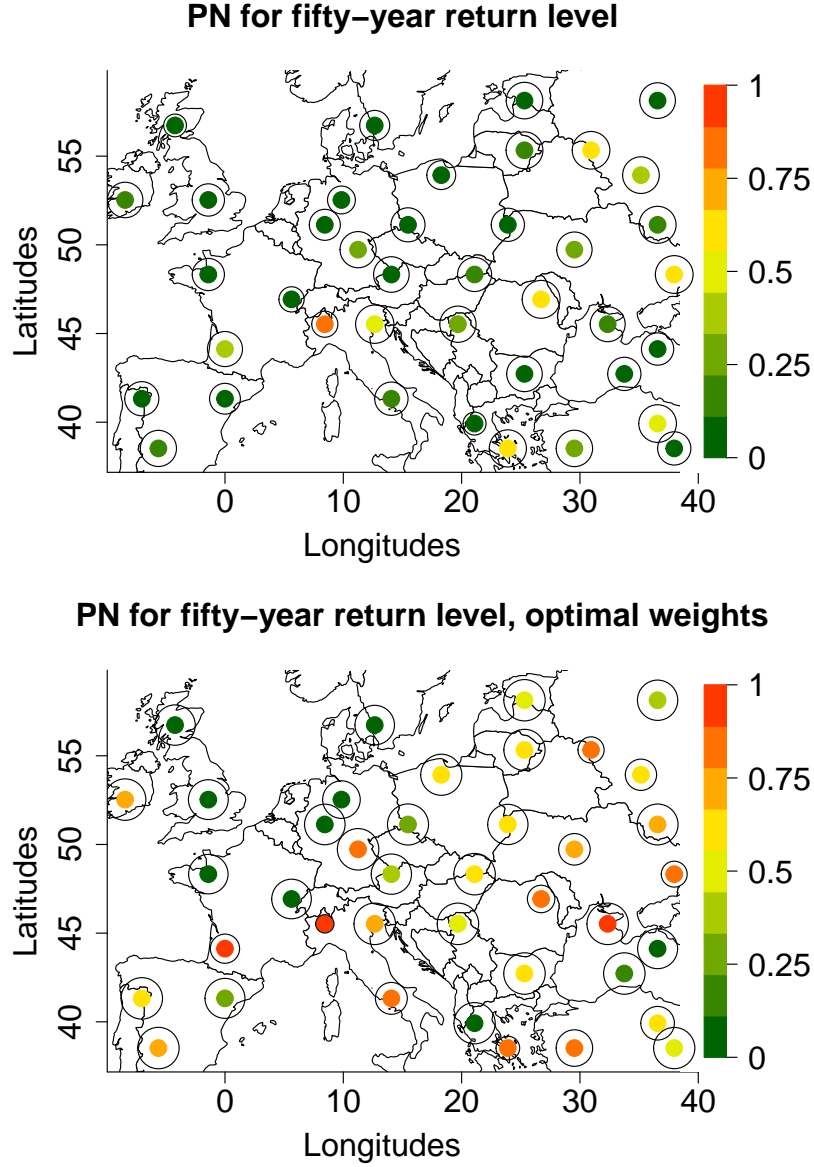


Fig 9: Necessary causation probabilities for a fifty-year return level of weighted maximal weekly winter precipitation in the counterfactual world for each cluster, calculated using equal weights (top) and optimal weights (bottom). The diameters of the black circles around the estimates are proportional to the length of bootstrap-based 95% confidence intervals.

might under- or overestimate the causation probabilities, while the proposed approach will take these changes into account. In addition, the application on heavy precipitation suggests that the multivariate approach can help in reducing the uncertainty of the estimates. Finally, the multivariate approach can highlight those grid points that maximise the probability of necessary causation through an adequate choice of the weights \mathbf{w} .

Some care is needed when applying the multivariate methodology. First of all, the model is restricted to homogeneous regions, since it assumes an equal shape parameter. This is a common assumption in multivariate EVT models. Moreover, the multivariate GPD is tailored for data that exhibit *asymptotic dependence*, i.e., the extremes in each grid point are expected to occur together. Finally, the dataset under study needs to be reasonably stationary in time. In future work, a non-stationary extension of the multivariate GPD could be proposed that would be highly relevant for longer periods of data from the factual world.

Another interesting research direction will be to extend the coupling between EVT and counterfactual theory to other types of extremes modeling in geosciences; see, e.g. [Hammerling, Katzfuss and Smith \(2017\)](#) or [Reich, Shaby and Cooley \(2014\)](#) for a Bayesian hierarchical point of view, [Shooter et al. \(2019\)](#) for asymptotic independence models, and [Ragone, Wouters and Bouchet \(2018\)](#) for rare event algorithms.

SUPPLEMENTARY MATERIAL

Supplement to “Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory”

(doi: [COMPLETED BY THE TYPESETTER](#); .pdf). The supplement includes results on the Gaussian MGPD model, probability weighted moment inference, a simulation study and further analysis for the precipitation data.

References.

- ANGÉLIL, O., STONE, D. and WEHNER, M. (2017). An Independent Assessment of Anthropogenic Attribution Statements for Recent Extreme Temperature and Rainfall Events. *Journal of Climate* **30** 5–16.
- BALKEMA, A. A. and DE HAAN, L. (1974). Residual life time at great age. *The Annals of Probability* **2** 792–804.
- BEIRLANT, J., GÖEGBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- BERNARD, E., NAVEAU, P., VRAC, M. and MESTRE, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate* **26** 7929–7937.
- CARREAU, J., NAVEAU, P. and NEPPEL, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research* **53** 4407–4426.

- CHEN, Y., MOUFOUMA-OKIA, W., MASSON-DELMOTTE, V., ZHAI, P. and PIRANI, A. (2018). Recent Progress and Emerging Topics on Weather and Climate Extremes Since the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Annual Review of Environment and Resources* **43** 35-59.
- COLES, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag Inc.
- COOLEY, D., HUNTER, B. D. and SMITH, R. L. (2017). *Handbook of Environmental and Ecological Statistics* Univariate and Multivariate Extremes for the Environmental Sciences. CRC Press.
- DAVISON, A. C. and HUSER, R. (2015). Statistics of Extremes. *Annual Review of Statistics and Its Application* **2** 203-235.
- DAVISON, A. C. and SMITH, R. L. (1990). Models for Exceedances over High Thresholds (with Comments). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **52** 393-442.
- DE FONDEVILLE, R. and DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105**.
- EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance* **33**. Applications of Mathematics, Springer-Verlag, Berlin.
- ENGELKE, S., DE FONDEVILLE, R. and OESTING, M. (2018). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika* **106** 127-144.
- FALK, M. and GUILLOU, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis* **99** 715-734.
- FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717-1737.
- FISCHER, E. M. and KNUTTI, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change* **5** 560-564.
- HAMMERLING, D., KATZFUSS, M. and SMITH, R. L. (2017). *Handbook of Environmental and Ecological Statistics* Climate Change Detection and Attribution. CRC Press.
- HANNART, A. and NAVEAU, P. (2018). Probabilities of Causation of Climate Changes. *Journal of Climate* **31** 5507-5524.
- HANNART, A., PEARL, J., OTTO, F., NAVEAU, P. and GHIL, M. (2016). Causal counterfactual theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society* **97** 99-110.
- HUSER, R., DAVISON, A. C. and GENTON, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes* **19** 79-103.
- KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in water resources* **25** 1287-1304.
- KATZFUSS, M., HAMMERLING, D. and SMITH, R. L. (2017). A Bayesian hierarchical model for climate change detection and attribution. *Geophys. Res. Lett.* **11** 5720-5728.
- KAUFMAN, L. and ROUSSEUW, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. Wiley Online Library.
- KEW, S. F., PHILIP, S. Y., JAN VAN OLDENBORGH, G., VAN DER SCHRIER, G., OTTO, F. E. L. and VAUTARD, R. (2019). The Exceptional Summer Heat Wave in Southern Europe 2017. *Bulletin of the American Meteorological Society* **100** S49-S53.
- KIM, Y.-H., MIN, S.-K., ZHANG, X., ZWIERS, F., ALEXANDER, L. V., DONAT, M. G. and TUNG, Y.-S. (2016). Attribution of extreme temperature changes during 1951-2010. *Climate dynamics* **46** 1769-1782.
- KING, A. D. (2017). Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future* **5** 1156-1168.

- KIRILIOUK, A. and NAVEAU, P. (2020). Supplement to ‘Climate extreme event attribution using multivariate peaks-over-thresholds modelling and counterfactual theory. DOI:.
- KIRILIOUK, A., SEGERS, J. and WARCHOL, M. (2016). Nonparametric estimation of extremal dependence. In *Extreme Value Modeling and Risk Analysis: Methods and Applications* CRC Press.
- KIRILIOUK, A., ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2019). Peaks over thresholds modelling with multivariate generalized Pareto distributions. *Technometrics* **61** 123–135.
- LUU, L. N., VAUTARD, R., YIOU, P., VAN OLDENBORGH, G. J. and LENDERINK, G. (2018). Attribution of extreme rainfall events in the South of France using EURO-CORDEX simulations. *Geophysical Research Letters* **45** 6242–6250.
- NATIONAL ACADEMIES OF SCIENCES, ENGINEERING AND MEDICINE (2016). *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, Washington, DC.
- NAVEAU, P., HANNART, A. and RIBES, A. (2020). Statistical Methods for Extreme Event Attribution in Climate Science. *Annual Review of Statistics and Its Application* **7** 89–110.
- OTTO, F. E., PHILIP, S., KEW, S., LI, S., KING, A. and CULLEN, H. (2018). Attributing high-impact extreme events across timescales—a case study of four different types of events. *Climatic change* **149** 399–412.
- PACIOREK, C. J., STONE, D. A. and WEHNER, M. F. (2018). Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather and Climate Extremes* **20** 69–80.
- PEARL, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- PICKANDS, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3** 119–131.
- RAGONE, F., WOUTERS, J. and BOUCHET, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences* **115** 24–29.
- REICH, B. J., SHABY, B. A. and COOLEY, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *Journal of Agricultural, Biological, and Environmental Statistics* **19** 119–135.
- ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018a). Multivariate peaks over thresholds models. *Extremes* **21** 115–145.
- ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018b). Multivariate generalized Pareto distributions: parametrizations, representations and properties. *Journal of Multivariate Analysis* **165** 117–131.
- ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930.
- SHABY, B. A. and REICH, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* **23** 638–48.
- SHOOTER, R., ROSS, E., TAWN, J. and JONATHAN, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics* **1** 1–18.
- STOTT, P. A., STONE, D. A. and ALLEN, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature* **432** 610.
- STOTT, P. A., CHRISTIDIS, N., OTTO, F. E., SUN, Y., VANDERLINDEN, J.-P., VAN OLDENBORGH, G. J., VAUTARD, R., VON STORCH, H., WALTON, P., YIOU, P. et al. (2016). Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Re-*

views: Climate Change **7** 23–41.

TAJVIDI, N. (1996). Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions, PhD thesis, Department of Mathematics, Chalmers, Göteborg.

ZSCHEISCHLER, J., WESTRA, S., VAN DEN HURK, B. J. J. M., SENEVIRATNE, S. I., WARD, P. J., PITMAN, A., AGHAKOUCHAK, A., BRESCH, D. N., LEONARD, M., WAHL, T. and ZHANG, X. (2018). Future climate risk from compound events. *Nature Climate Change* **8** 469–477.